

# Reimagining Peer Review: A Novel Framework for Evaluating Scientific Discoveries

A team of immunologists designed a study (see it on [BioRxiv](#)) to test a new strategy for improving the peer review process and strengthening the rigor and reliability of scientific publications. While scientific publishing originally revolutionized how knowledge is shared, its modern form often incentivizes overstating results to maximize perceived impact rather than ensuring rigor. The Discovery Stack model we tested seeks to realign peer review with its core purpose of upholding scientific rigor.

## The Problem

Peer review is the cornerstone of scientific quality control. It shapes how discoveries are generated, communicated, applied, and valued. Yet, the current journal-centric publishing system does not effectively serve the scientific community or the public. It is hindered by long timelines, reviewer bias, high costs, editorial gatekeeping, and misaligned measures of impact.

As practiced now, peer review often conflates two distinct dimensions of evaluation: **Quality** and **Impact**. **Quality** reflects how well a study's conclusions are supported by rigorous, reproducible data, while **Impact** reflects the extent to which the findings advance understanding, fill knowledge gaps, and guide future research. When these dimensions are blurred, high-impact findings can mask weak evidence, while rigorous but incremental work is undervalued. This misalignment favors novelty over rigor and undermines reproducibility.

## The Hypothesis

We hypothesized the Discovery Stack model would **facilitate separate evaluation of a manuscript's Quality and Impact** using standardized, structured review criteria. Additionally, we posited that the **use of a line-editing method for reviewer-author interaction** would foster a "peer-improvement" mindset, encouraging reviewers to provide clear, constructive, and actionable feedback that strengthens the rigor of the study.

## The Discovery Stack Pilot

To test this hypothesis, we launched the Discovery Stack Pilot to evaluate the feasibility and effectiveness of a structured, multi-phase review model that uses standardized metrics and an in-line annotation tool to first evaluate *Quality* and then *Impact* (**Fig. 1**).

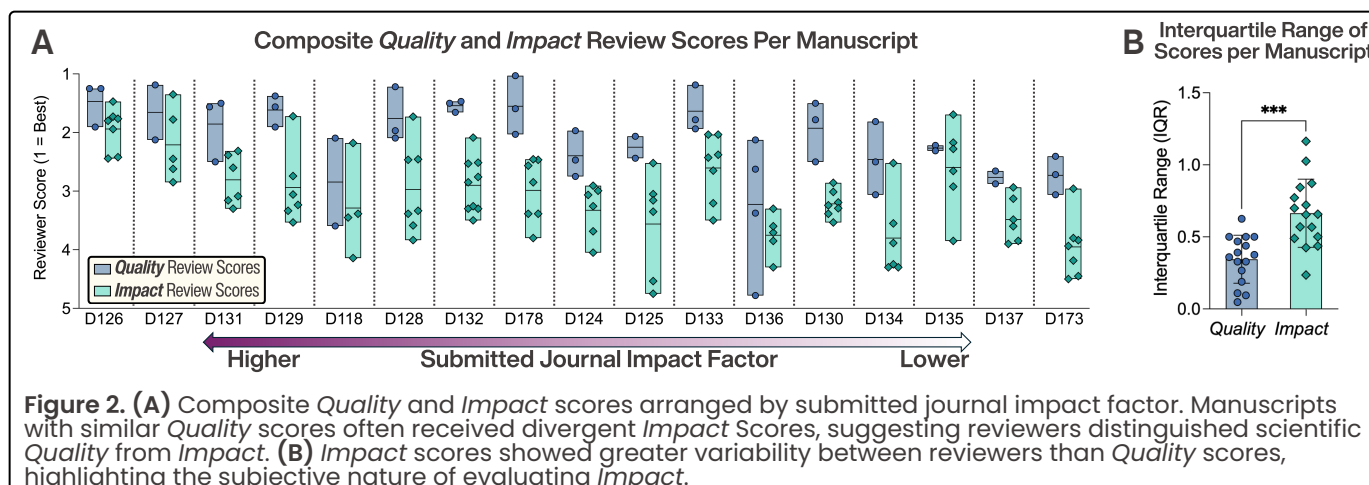
Manuscripts enrolled in the pilot were posted on bioRxiv and reviewed in parallel through traditional journal processes, allowing for direct comparison between the systems. A total of 162 reviews were completed, and survey data collected from 86 participants.



**Figure 1. The Discovery Stack Model.** The model includes three sequential phases: **(1) Quality Review** – Reviewers evaluated the rigor of the study, including methodology, experimental design, statistical analyses, and whether the data supported the conclusions. **(2) Author Response**, and **(3) Impact Review** – Reviewers assessed the extent the findings advanced scientific understanding, filled knowledge gaps, and influenced multiple fields.

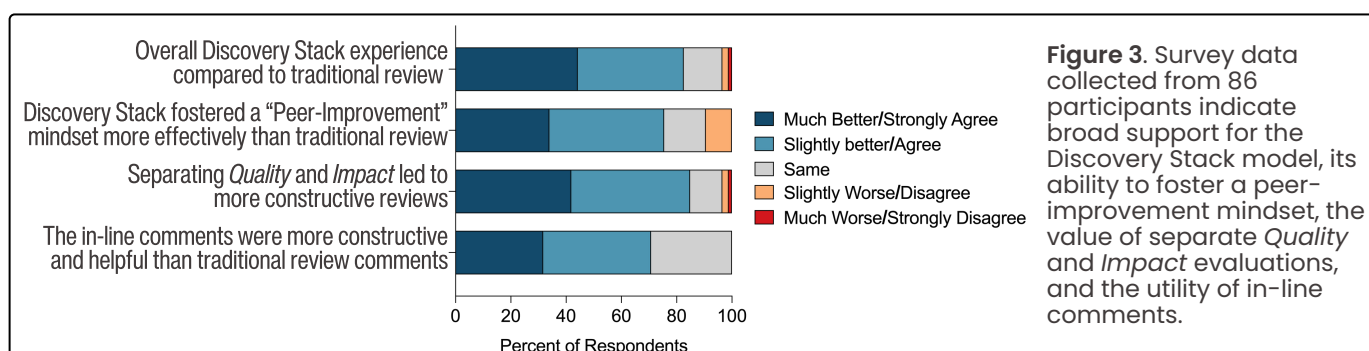
## Key Finding: Reviewers Clearly Distinguished *Quality* from *Impact*

The Discovery Stack model effectively enabled reviewers to evaluate *Quality* and *Impact* as separate dimensions. *Quality* scores were more consistent across reviewers, whereas *Impact* scores showed greater variability, reflecting their more subjective nature (Fig. 2).



## Key Finding: Strong Support for Core Elements of Discovery Stack

Authors and reviewers expressed strong support for core elements of the Discovery Stack model. Participants reported that the process fostered a peer-improvement mindset, and that separating *Quality* and *Impact* evaluations led to clearer, more constructive reviews. In addition, the in-line annotation tool improved review clarity, efficiency, and collegiality (Fig. 3).



## Additional Observations

The study also revealed broad support for standardized metrics for evaluating both *Quality* and *Impact*. In future iterations, these metrics could help researchers search for and prioritize relevant research. Reviewer identity disclosure was associated with greater perceived transparency and higher *Impact* scores, suggesting reviewers may be more likely to identify themselves when giving favorable evaluations. Finally, familiarity with the project and personal outreach markedly improved reviewer acceptance rates.

## Conclusions

By separating *Quality* from *Impact*, introducing standardized metrics, and fostering a peer-improvement culture, the Discovery Stack framework offers a clear path toward transparent, equitable, and evidence-driven peer review focused on enhancing scientific rigor.

Beyond improving how individual manuscripts are evaluated, the Discovery Stack model establishes the foundation for a new platform to curate peer-reviewed and validated research, with the potential to accelerate the dissemination of discoveries and enable the scientific community to identify and build upon rigorous, impactful work.